

# Prediction of Halocarbon Toxicity from Structure: A Hierarchical QSAR Approach

*B.D. Gute, K. Balasubramanian, K. Geiss, S.C. Basak*

This article was submitted to  
International Conference on Chemical Mixtures  
Atlanta, GA  
September 10-12, 2002

**April 11, 2003**

*U.S. Department of Energy*

Lawrence  
Livermore  
National  
Laboratory

## DISCLAIMER

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

This is a preprint of a paper intended for publication in a journal or proceedings. Since changes may be made before publication, this preprint is made available with the understanding that it will not be cited or reproduced without the permission of the author.

This report has been reproduced directly from the best available copy.

Available electronically at <http://www.doc.gov/bridge>

Available for a processing fee to U.S. Department of Energy  
And its contractors in paper from  
U.S. Department of Energy  
Office of Scientific and Technical Information  
P.O. Box 62  
Oak Ridge, TN 37831-0062  
Telephone: (865) 576-8401  
Facsimile: (865) 576-5728  
E-mail: [reports@adonis.osti.gov](mailto:reports@adonis.osti.gov)

Available for the sale to the public from  
U.S. Department of Commerce  
National Technical Information Service  
5285 Port Royal Road  
Springfield, VA 22161  
Telephone: (800) 553-6847  
Facsimile: (703) 605-6900  
E-mail: [orders@ntis.fedworld.gov](mailto:orders@ntis.fedworld.gov)  
Online ordering: <http://www.ntis.gov/ordering.htm>

OR

Lawrence Livermore National Laboratory  
Technical Information Department's Digital Library  
<http://www.llnl.gov/tid/Library.html>

# **Prediction of Halocarbon Toxicity from Structure:**

## **A Hierarchical QSAR Approach**

Brian D. Gute,<sup>a</sup> K. Balasubramanian,<sup>b</sup> K. Geiss<sup>c</sup>, and S.C. Basak<sup>a,1</sup>

<sup>a</sup>Natural Resources Research Institute, University of Minnesota Duluth,  
5013 Miller Trunk Hwy., Duluth, MN 55811

<sup>b</sup>Department of Applied Science, University of California Davis, Livermore, CA 94550;  
Chemistry & Material Science Directorate, Lawrence Livermore National Laboratory, University  
of California PO Box 808, L-268, Livermore, CA 94550 and Glenn T. Seaborg Center,  
Lawrence Berkeley National Laboratory, Berkeley CA 94720

<sup>c</sup>Operational Toxicology Branch, Air Force Research Laboratory,  
Wright-Patterson AFB, OH 45433

---

<sup>1</sup> Dr. Subhash C. Basak  
Natural Resources Research Institute  
5013 Miller Trunk Hwy.  
Duluth, MN 55811

T—(218)720-4230  
F—(218)720-4328  
E—sbasak@nrri.umn.edu

## **Abstract**

Mathematical structural invariants and quantum theoretical descriptors have been used extensively in quantitative structure-activity relationships (QSARs) for the estimation of pharmaceutical activities, biological properties, physicochemical properties, and the toxicities of chemicals. Recently our research team has explored the relative importance of various levels of chemodescriptors, i.e., topostructural, topochemical, geometrical, and quantum theoretical descriptors, in property estimation. This study examines the contribution of chemodescriptors ranging from topostructural to quantum theoretic calculations up to the Gaussian STO-3G level in the prediction of the toxicity of a set of twenty halocarbons. We also report the results of experimental cell-level toxicity studies on these twenty halocarbons to validate our models.

## **Keywords**

## 1. Introduction

Halocarbons constitute an important group of industrial chemicals being widely used as solvents and synthetic intermediates in large quantities around the world (Fishbein, 1985). Chemicals of this class have been found to be mutagenic, carcinogenic, and strongly hepatotoxic. The high volume of use of these chemicals indicates that there is a great potential for the release of this class of chemicals and resulting adverse effects on human and environmental health from exposure to them. Therefore, there is an interest in assessing the potential toxic effects of these chemicals.

Hepatotoxicity of chemicals can conveniently assayed *in vitro* using cultured hepatocytes as compared to whole animals. Once a well-designed subset of chemicals is tested, one can develop quantitative structure-activity relationship (QSAR) models to understand the structural basis of toxicity; a good-quality QSAR can also help in estimating the potential toxicity of untested chemical of the same class.

QSAR models can be formulated based on experimentally derived descriptors, *viz.* steric or electronic parameters derived from physical organic chemistry (Hansch and Leo, 1995), hydrogen bond donor acidity, hydrogen bond acceptor basicity, experimental hydrophobicity, or parameters which can be computed from molecular structure without the input of experimental data. Whereas the experimentally based QSARs work well with narrow classes of chemicals, such parameters are not available for diverse groups of chemicals. Another advantage of theoretically calculated descriptors is that they are available for any molecule, real or hypothetical. Therefore, the latter group of descriptors can be used in the *in silico* evaluation of compounds not yet synthesized.

Our research team has been involved in the use of calculated molecular descriptors in developing QSAR models. We have recently formulated a hierarchical QSAR (HiQSAR) approach where parameters of increasing chemical or computational complexity are used in a graduated manner. In particular, we have used topostructural (TS), topochemical (TC), geometrical (3-D), and quantum chemical (semiempirical as well as *ab initio*) parameters in QSAR model development for various properties including the complement inhibitory properties of benzamidines (Basak et al., 1999), the mutagenicity of a set of 95 aromatic amines (Basak et al., 1998), the skin penetration profiles for 60 polycyclic aromatic hydrocarbons (Gute et al., 1999), and the toxicity of a group of 55 halocarbons to *aspergillus niger* (Basak et al., 2003). In this paper we have used a variant of the HiQSAR approach to develop predictive models for the toxicity of a group of 20 halocarbons to primary hepatocytes. This variant approach was implemented due to the small size of the chemical data set.

## **2. Materials and Methods**

Presented below are the methods and materials used in the isolation of rat F-344 hepatocytes, exposure of the cells to the toxicants, and the determination of cell-level toxicity for the halocarbons after exposure. Scientists at Wright Patterson Air Force Base generated the data used in this study.

### *2.1 Experimental Methods*

#### *2.1.1 Chemicals*

Chemicals and biological reagents were purchased from Sigma Chemical Co. (St. Louis, MO) unless otherwise stated. This includes most of the halogenated test chemicals (CHBrCl<sub>2</sub>,

CHBr<sub>2</sub>Cl, CHBr<sub>3</sub>, CH<sub>2</sub>Br<sub>2</sub>, CH<sub>2</sub>BrCl, CH<sub>2</sub>Cl<sub>2</sub>, CCl<sub>4</sub>, CBrCl<sub>3</sub>, CBr<sub>2</sub>Cl<sub>2</sub>, CBr<sub>4</sub>, C<sub>2</sub>HCl<sub>3</sub>, 1,1,2,2-C<sub>2</sub>H<sub>2</sub>Cl<sub>4</sub>, 1,2-C<sub>2</sub>H<sub>4</sub>Cl<sub>2</sub>, 1,2-C<sub>2</sub>H<sub>4</sub>BrCl, 1,1,2-C<sub>2</sub>H<sub>3</sub>Br<sub>3</sub>, 1,2-C<sub>2</sub>H<sub>4</sub>Br<sub>2</sub>, 1,1,2-C<sub>2</sub>H<sub>3</sub>Cl<sub>3</sub>, 1,1,1-C<sub>2</sub>H<sub>3</sub>Cl<sub>3</sub>). C<sub>2</sub>Cl<sub>4</sub> and CHCl<sub>3</sub> were purchased from Fisher Scientific (Pittsburgh, PA). Chemicals were used as provided by the manufacturer, without further purification. The halogenated chemicals are summarized in Table 1, shown with their CAS number, catalog number, purity/grade, an indication of which test set they were used in (preliminary model set, P; full model set, M; evaluation set, E), and their experimental dose ranges. Ketamine (Injectable, U.S.P. grade) was purchased from Parke-Davis (Morris Plains, NJ). Xylazine (Injectable, U.S.P. grade) was purchased from Mobay Corporation (Shawnee, KS). Collagenase (Type D) and Protein Assay ESL reagents were purchased from Roche (formerly Boehringer-Mannheim) Biochemicals (Indianapolis, IN). Type I rat tail collagen was purchased from Upstate Biotechnology (Lake Placid, NY). CHEE's modified Eagle Medium (Formula No. 88-5046EA) and Hank's balanced salt solution (HBSS) were purchased from GibcoBRL/Life Technologies (Rockville, MD).

### 2.1.2 Animals

Fischer CD<sup>+2</sup>F<sup>R</sup>(F344)/CrIBR (F-344) male rats (225-300 g) were obtained from Charles River Laboratories. All animals used in this study were handled in accordance with the principles stated in the “*Guide for the Care and Use of Laboratory Animals*,” National Research Council, 1996, and the Animal Welfare Act of 1966, as amended. Rats were anesthetized with 1 mL/kg of a mixture of ketamine (70 mg/mL) and xylazine (6 mg/mL) prior to undergoing *in situ* liver perfusion.

### *2.1.3 Hepatocyte Isolation and Culture*

Rat livers were digested by perfusion using a two-step method. In the first perfusion step, the liver was perfused via the hepatic portal vein with perfusion buffer (37°C) consisting of Hank's balanced salt solution (HBSS; pH 7.2) lacking calcium and magnesium and supplemented with 15 mM 4-(2-hydroxyethyl)-1-piperazineethane-sulfonic acid (HEPES), heparin (2.0 U/mL) and ethylene-bis(oxyethylenenitrilo)-tetraacetic acid (EGTA; 0.5 mM). This first step of the antigrade two-step perfusion method was accomplished by canulation of the portal vein followed by clamping of the posterior vena cava anterior to the diaphragm. Flow of perfusion buffer was then activated (20 mL/min), immediately followed by the cutting of the posterior vena cava (anterior to the renal vein) to allow drainage of the perfusion buffer. Following complete removal of blood from the liver, the second step of the liver perfusion was initiated by continuous perfusion with digestion buffer. In this second perfusion step, digestion buffer (37°C) consisting of complete HBSS (pH 7.2) supplemented with collagenase (0.26 Wunsche Units/mL), was perfused through the liver to digest interstitial connective tissue. Perfusion was continued until hepatocytes were completely disaggregated (approximately 20 min). Viable primary rat hepatocytes were washed (37°C) with complete HBSS (pH 7.2) and enriched three times by low speed centrifugation at 50 g for 3 min. Typical viabilities of isolated hepatocytes ranged from 80 to 95% with yields of 250 to 400 million cells as determined by trypan blue dye exclusion. For cell culture studies, freshly isolated hepatocytes in suspension were adjusted to a cell density of  $1.0 \times 10^6$  cell/mL in cell attachment medium consisting of CHEEs modified culture medium (CHEE; pH 7.2) supplemented with HEPES (10 mM), insulin/transferrin/sodium selenite solution (final concentrations of 5 µg/mL, 5 µg/mL and 5 ng/mL, respectively), gentamycin (0.1 mg/mL), and dexamethasone (0.4 mg/mL). Cells were seeded in 6-well ( $1.0 \times$



10<sup>6</sup> cells/well) culture plates. Plates were previously coated with Type I rat tail collagen (25.0 µg/mL stock) at 2.6 µg/cm<sup>2</sup>. After 4 h of incubation in a 95% air/5% CO<sub>2</sub> incubator at 37°C, cell attachment medium was removed and rat hepatocytes were incubated in fresh CHEEs culture medium lacking dexamethasone. Hepatocytes were incubated for an additional 20 h in order to recover from the stress incurred during the isolation procedure. Exposures to halogenated chemicals were initiated at this point. The overall cell culture scheme is shown in Figure 1.

#### *2.1.4 Cell Culture and Chemical Exposure*

All exposures to the test chemicals were accomplished in the VITROBOX™ exposure system (Geiss and Frazier, 2000). The system was basically a rectangular glass chamber, approximately 8x25x25 cm interior dimensions (~ 5L volume). For each chemical dose, there was a separate VITROBOX™ chamber and Tedlar dosing bag. Chemical dose ranges are also shown in Table 1.

Prior to exposures (~ 30 min) the appropriate chemical concentration was added to the Tedlar dosing bags and placed back into a 37°C incubator to allow the chemical to volatilize. Also, immediately prior to cell culture dosing, the chemicals were diluted into the appropriate medium. No other solvents or vehicles (e.g. ethanol) were used in preparation of the dosing solutions. Media in the cell culture plates was replaced with the chemical dosing media. Plates were then immediately placed into the chamber without plate lids. The chamber was then closed and the dosing and capture bags were attached to the chamber. The dosing of the chamber was accomplished by manually compressing the dosing bag, forcing the dosing atmosphere into the chamber, while it exhausted into the capture bag. Following the dosing process, all ports were

sealed and the chamber was placed into a 37°C incubator. The cells were exposed in the chambers for 4 h and then removed from the chambers for analysis.

### *2.1.5 Biochemical Assays*

Six assays were used to determine cytotoxicity and oxidative stress. MTT and LDH are indicators of cellular viability, i.e. mitochondrial function and membrane integrity, respectively. Reactive oxygen species (ROS) were measured as indicators of reactive oxidant molecules that are produced in the cell and result in oxidant damage to the cell. Lipid peroxidation (TBARS) was measured as an indicator of oxidative damage to cellular lipid molecules. Catalase (CAT) activity was determined as a representative antioxidant enzyme. Decreases in catalase activity can contribute to an increased oxidative stress situation in the cell. Total cellular thiols (SH) were determined to assess the antioxidant thiol status.

## *2.2 Theoretical Methods*

### *2.2.1 Calculation and Classification of Molecular Descriptors*

A complete list of the 187 molecular descriptors used in the current study is provided in Table 2. In Table 2, these parameters are partitioned into classes ordered hierarchically based on theoretical complexity. At the lowest level of the hierarchy are the topostructural indices that encode information strictly about the adjacency and topological distances of atoms in molecular structures. The topochemical indices are more complex, encoding information not only about the topology but also chemical properties of the molecular structures. The geometrical indices are more complex still as they encode information about the shape and 3-dimensional aspects of the

structure. Finally, the quantum chemical indices are at the upper end of the hierarchy, encoding electronic information. The software programs used for the calculation of these descriptors include POLLY v2.3 (Basak et al., 1988), Triplet (Filip et al., 1987), Molconn-Z v3.50 (Molconn-Z v3.50, 2000), Sybyl v6.4 (Sybyl v6.4, 1997), MOPAC v6.00 (Stewart, 1990), and Gaussian 98W (Gaussian 98, 2001).

POLLY v2.3 and Triplet each calculate approximately 100 topostructural and topochemical descriptors, collectively known as topological indices, including the Wiener number (Wiener, 1947), molecular connectivity indices developed by Randić (1975) and Kier and Hall (1986), frequency of path lengths of varying size (Kier and Hall, 1986), information theoretic indices defined on distance matrices of graphs using the methods of Bonchev and Trinajstić (1977), Basak et al., Roy et al., and those of Raychaudhury *et al.* (Basak et al., 1980), and parameters defined on the neighborhood complexity of vertices in hydrogen-filled molecular graphs (Roy et al., 1984) as well as the triplet indices (Filip, 1987). The triplets result from a matrix, main diagonal column vector, and free term column vector which are converted into a system of linear equations. The notation used to represent the vectors and matrices is as follows:

A = Adjacency matrix

V = Vertex degree

S = Distance sum

N = Total number of vertices in the graph

Z = Atomic number

D = Distance matrix

1 = Unity matrix.

After the system of  $N$  linear equations is solved, the local vertex invariants,  $x_i$ , are assembled into a triplet descriptor based on one of the following operations:

1. Summation,  $E_i x_i$ ;
2. Summation of squares,  $E_i x_i^2$ ;
3. Summation of square roots,  $E_i x_i^{1/2}$ ;
4. Sum of inverse square root of cross-product over edges  $ij$ ,  $E_{ij}(x_i x_j)^{-1/2}$ ;
5. Product,  $N(E_i x_i)^{1/N}$ .

From Molconn-Z v3.50, we obtain approximately 150 additional descriptors, including an extended set of molecular connectivity indices, electrotopological state descriptors, general polarity descriptors, hydrogen bonding descriptors, and Kappa shape indices. Sybyl (Sybyl v6.4, 1997) and MOPAC (Stewart, 1990) calculate a relatively small number of geometrical and quantum chemical descriptors, respectively. With respect to the geometrical descriptors generated by Sybyl, Concord 3.2.1 (Concord v3.2.1, 1997) was used to determine the 3-D coordinates of the atoms. The Concord geometries were then used in Gaussian98W to generate the *ab initio* STO-3G basis set calculations. We are also in the process of using more intensive computations such as coupled cluster methods that use much larger basis sets.

< Table 2 about here >

### *2.3 Data Preparation and Statistical Analysis*

A total of 379 molecular descriptors were calculated for each of the chemicals in this set of halocarbons. However, many were removed from the final descriptor set. Any descriptor with a value of zero for all twenty compounds was eliminated from use in the subsequent analysis. Also, triplet descriptors that could not be calculated for all twenty compounds were removed, a limitation of the Triplet program is that it cannot calculate descriptors for molecules with fewer than three non-hydrogen atoms. In addition, the CORR procedure of the SAS statistical software package (SAS Institute, 1988) was used to identify perfectly correlated descriptors, i.e. those descriptor pairs with a correlation coefficient equal to 1. In each case, only one of the perfectly correlated descriptors was retained and used in the subsequent analysis. At this point, 187 molecular descriptors remained. Prior to analysis, the descriptors were transformed by the natural logarithm due to the fact that their scales differed by several orders of magnitude. Appropriate constants were first added to any descriptors with values less than or equal to zero before performing the transformation.

Due to the large number of independent variables available (187) and the small number of instances, twenty individual chemicals, a modified form of the hierarchical method was employed in this study. The set of 174 topological and geometric indices was clustered, using the SAS procedure VARCLUS (SAS Institute, 1988), into six, essentially unidimensional, disjoint clusters. To keep the number of independent variables selected for modeling to a minimum, the variable clustering procedure was forced to produce only six clusters and the one index most correlated with each cluster was selected for modeling. The quantum chemical indices were taken as two distinct sets, not to be combined, since they represent the same calculations at different levels of quantum theory. Since each of these sets contained no more than seven

indices, we retained all indices for modeling. While starting with six or seven indices to model a set of twenty chemicals may seem to be an overly-large number of independent variables, six indices is the cut-off for using thirty-percent as many independent variables as instances, a common guideline for selecting the number of variables to be used. Additionally, we felt that using seven semi-empirical indices was a better choice than trying to decide which one to remove. Each toxicological endpoint was modeled using the set of six topological indices selected by variable clustering, the set of seven semi-empirical indices, and the set of six *ab initio* indices. Finally, the topological indices selected for optimal modeling of each property was combined, separately, with the semi-empirical and *ab initio* indices providing the best models. In this manner, we kept the number of independent variables used to, at most, six or seven indices per model.

Regression modeling was accomplished using the SAS procedure REG (SAS Institute, 1988) on three distinct sets of indices for each of the six dependent variables (toxicological endpoints). These sets were constructed as part of the HiQSAR approach to model development. In this case, the hierarchy began with the simpler topological parameters, including the topostructural, topochemical, and geometrical indices. After using the topological indices to model the activity, the quantum chemical parameters were used in modeling. From these three sets of models, we combined the topological indices selected in the best models with the semi-empirical and *ab initio* indices for the top two tiers of the hierarchy.

<Insert Table 3 here>

### 3. Results

Variable clustering on the set of 174 topostructural, topochemical, and geometrical indices resulted in the retention of six indices:  ${}^1\chi^v$ ,  ${}^5\chi^b$ , ANN<sub>5</sub>, ANV<sub>5</sub>, nclass, and SHCsats. Two of these are topostructural indices (ANN<sub>5</sub> and ANV<sub>5</sub>) while the remaining four are topochemical indices. None of the geometrical indices were selected by variable clustering. Definitions for these indices can be found in Table 2.

Of the models developed to estimate mitochondrial function, MTT, the addition of the AM1 semi-empirical parameters to the topological indices resulted in the most reliable model among the models considered here. We expect further improvement through the use of state-of-the-art coupled cluster methods in conjunction with large basis sets. As can be seen in Table 3, the models using only topological or quantum chemical indices do a poor job of modeling MTT. While the topological and semi-empirical models explain approximately 62% of the variance in the data, their cross-validated  $R^2$  show that these models are not robust. The combination of topological and semi-empirical indices, resulting in an  $R^2$  of 71%, shows significant improvement over the models using individual classes of indices, though modeling using the *ab initio* parameters results in the selection of the same topological-only model. This could be attributed in part due to the limited STO-3G basis set, and we expect better results using larger basis sets and coupled cluster methods. Therefore, the combination of topological indices (ANN<sub>5</sub> and  ${}^1\chi^v$ ) and AM1 heat of formation ( $\Delta H_f$ ) results in the most robust model for MTT, and the model equation is presented below:

$$1/\text{MTT} = -12.3719 + 5.6850 (\text{ANN}_5) + 2.8142 ({}^1\chi^v) + 0.8388 (\Delta H_f) \quad \text{Eq. 1}$$

$$n = 20, R^2 = 0.7115, F = 13.15, s.e. = 1.01$$

Modeling membrane integrity, LDH, demonstrated results similar to the modeling of mitochondrial function. However, in this case the models were all significantly more robust. The topological modeling provides a reasonable three-parameter ( $\text{ANN}_5$ ,  ${}^1\chi^v$ , and SHCsats) QSAR model that is nearly as robust as the four-parameter model incorporating the three topological indices and the AM1 parameter,  $E_{\text{HOMO}}$ . Neither of the quantum chemical-only models were significant as compared with the models containing topological indices. Equation 2 presents the topological model, deemed to be the better model for its use of fewer independent variables while maintaining a level of robustness nearly equal to that of the model combining topological and semi-empirical parameters.

$$\begin{aligned} \text{Ln(LDH)} &= 11.8701 - 4.1918 (\text{ANN}_5) - 4.6102 ({}^1\chi^v) + 2.0897 (\text{SHCsats}) & \text{Eq. 2} \\ n &= 20, R^2 = 0.7848, F = 19.44, s.e. = 0.85 \end{aligned}$$

In predicting the level of reactive oxygen species (ROS) produced by exposure to halocarbons, it can be seen once again that the topological indices outperform the quantum chemical indices. The two quantum chemical models show extremely poor model reliability when cross-validation is attempted (see Table 3), while the four-parameter topological model ( $\text{ANN}_5$ ,  ${}^1\chi^v$ ,  $n_{\text{class}}$ , and  ${}^5\chi^b_{\text{C}}$ ) is robust and shows good predictive capability. However, with the addition of the *ab initio* parameters to the topological indices, a model is found that only three indices ( ${}^1\chi^v$ ,  ${}^5\chi^b_{\text{C}}$ , and  $E_{\text{HOMO-1}}$ ) while maintaining model robustness and accuracy. This model is presented below (Eq. 3):

$$\text{Ln(ROS)} = 28.9247 - 14.2212 ({}^1\chi^v) + 8.5120 ({}^5\chi^b_{\text{C}}) + 22.6687 (E_{\text{HOMO-1}}) \quad \text{Eq. 3}$$



$$n = 20, R^2 = 0.7966, F = 20.88, s.e. = 0.92$$

In modeling the remaining three toxicity endpoints, we see similar results to modeling the formation of reactive oxygen species. The topological indices and the combination of topological and *ab initio* parameters perform nearly identically, while the use of quantum chemical indices alone do not come close to the performance of the other two models. In predicting lipid peroxidation (TBARS), we see nearly identical topological and topological plus *ab initio* models. Both models use only three parameters; ANN<sub>5</sub>,  $^1\chi^v$ , and  $^5\chi^b_C$ , and ANN<sub>5</sub>,  $^5\chi^b_C$ , and E<sub>HOMO-1</sub>; respectively, and their other regression parameters are nearly identical. Therefore, since the topological indices are much less “expensive” to calculate, in terms of computational resources and time, the topological model has been selected as the “better” model and is presented below in equation 4.

$$\text{Ln(TBARS)} = 13.7701 - 5.7991 (\text{ANN}_5) - 4.3134 (^1\chi^v) + 9.7829 (^5\chi^b_C) \quad \text{Eq. 4}$$

$$n = 20, R^2 = 0.7402, F = 15.19, s.e. = 0.85$$

In the case of estimating total cellular thiol levels, the four-parameter model using both topological and *ab initio* parameters (Eq. 5) was selected over the four-parameter model using only topological indices since it showed a 3% increase in cross-validated  $R^2$  and a drop in standard error of 0.04 natural logarithm units.

$$\begin{aligned} \text{Ln(Thiols)} = & 6.8244 - 12.1873 (\text{ANN}_5) + 3.2577 (\text{nclass}) \\ & + 14.9063 (^5\chi^b_C) - 12.5730 (\text{E}_{\text{HOMO-1}}) \end{aligned} \quad \text{Eq. 5}$$

$$n = 20, R^2 = 0.7976, F = 14.78, s.e. = 0.90$$

Finally, in estimating catalase activity, the topological only model was again selected as the simplest model. Just as in the case of the lipid peroxidation model, the addition of the *ab initio*, quantum chemical parameters resulted in a nearly identical model to that produced using the simpler topological indices. There is no real distinguishable difference between the two models. The topological model estimating catalase activity is presented as equation 6.

$$\begin{aligned} \text{Ln(Catalase)} = & 13.4695 - 5.2477 (\text{ANN}_5) + 5.4479 (\text{ANV}_5) - 3.7743 ({}^1\chi^v) \\ & - 2.2602 (\text{nclass}) \end{aligned} \quad \text{Eq. 6}$$

$$n = 20, R^2 = 0.8167, F = 16.71, s.e. = 0.83$$

Results of QSAR studies reported in this paper mentioned above and summarized in table 3 show that the models formulated using the topological descriptors usually performed better than models developed using quantum chemical descriptors alone. Some improvement can be achieved by combining topological and quantum chemical parameters, but the only truly significant improvement in modeling was seen for the estimation of mitochondrial function (MTT). More refined techniques with larger basis sets are needed to check if MTT will be impacted by quantum chemical parameters.

#### 4. Discussion

Graph theoretical (topological) indices encode information regarding the shape, size, branching, and cyclicity associated with molecular structure (Trinajstić, 1992). They have also been found

to be correlated with properties such as the hydrophobicity ( $\log P$ ) of molecules that plays a crucial role in determining the entry of chemicals into biological systems. It is possible that the topological indices, by representing the general structural features, are capable of explaining a large portion of the variance in the hepatotoxicity of the halocarbons. On the other hand, under anaerobic conditions liver halocarbon toxicity is deemed to be strongly determined by dissociative electron attachment. If such a mechanism is responsible for the hepatotoxicity of the halocarbons studied in this paper, calculated indices such as vertical electron affinity (VEA) should improve model quality over and above the ability of the topological indices. However computations of electron affinities require intensive level of computations that include large basis sets and methods such as coupled cluster singles + doubles with triple correction or CCSD(T) since the electron attachment process involves the lowest unoccupied molecular orbital (LUMO) which is not adequately described by lower order methods. Such studies are being carried out and will be reported subsequently.

### **Acknowledgements**

This is contribution number XXX from the Center for Water and the Environment of the Natural Resources Research Institute. Research reported in this paper was supported in part by grant F49620-02-1-0138 from the United States Air Force Office of Scientific Research. The work at UC Davis was supported in part by National Science Foundation grant CHE-0236434 while the work at LLNL was performed under the auspices of US Department of Energy by the University of California under Contract No.W-7405-Eng-48

### **References**

- Basak, S.C., Balasubramanian, K., Gute, B.D., Mills, D., 2003. Prediction of cellular toxicity of halocarbons from their computed chemodescriptors: A hierarchical QSAR approach, J. Chem. Inf. Comput. Sci., submitted.
- Basak, S.C., Gute, B.D., Ghatak, S., 1999. Prediction of complement inhibitory activity of benzamidines using topological and geometrical parameters. J. Chem. Inf. Comput. Sci. 39, 255-260.
- Basak, S.C., Gute, B.D., Grunwald, G.D., 1998. Relative effectiveness of topological, geometrical, and quantum chemical parameters in estimating mutagenicity of chemicals. In: Chen, F., Schuurmann, G., Eds., Quantitative Structure-activity Relationships in Environmental Sciences VII, SETAC Press, Pensacola, FL., pp. 245-261; Basak, S.C., Mills, D.R., Balaban, A.T., Gute, B.D., 2001. Prediction of mutagenicity of aromatic and heteroaromatic amines from structure: A hierarchical QSAR approach, J. Chem. Inf. Comput. Sci. 41, 671-678.
- Basak, S.C., Harriss, D.K., Magnuson, V.R., 1988. POLLY v2.3, Copyright of the University of Minnesota.
- Basak, S.C., Roy, A.B., Ghosh, J.J., 1980. In: Avula, X.J.R., Bellman, R., Luke, Y.L., Rigler, A.K., Eds., Proceedings of the Second International Conference on Mathematical Modelling, University of Missouri-Rolla, Rolla, Missouri; Roy, A.B., Basak, S.C., Harriss, D.K., Magnuson, V.R., 1984. In: Avula, X.J.R., Kalman, R.E., Lapis, A.I., Rodin, E.Y., Eds., Mathematical Modelling in Science and Technology, Pergamon Press, New York; Raychaudhury, C., Ray, S.K., Ghosh, J.J., Roy, A.B., Basak, S.C., J. Comput. Chem. 5.
- Bonchev, D., Trinajstić, N., 1977. J. Chem. Phys. 67, 4517-4533.

Concord v3.2.1, 1997. Tripos Associates, Inc.: St. Louis, MO.

Filip, P.A., Balaban, T.S., Balaban, A.T., 1987. *J. Math. Chem.* 1, 61-83; Basak, S.C., Balaban, A.T., Grunwald, G.D., Gute, B.D., 2000. *J. Chem. Inf. Comput. Sci.* 40, 891-898.

Fishbein, L., 1985. In: Fishbein, L., O'Neill, I. K., Eds., *Environmental Carcinogens - Selected Methods of Analysis*. International Agency for Research on Cancer, Lyon, pp. 47-67; Pearson, C. R., 1982. In: Hutzinger, O., Ed., *Handbook of Environmental Chemistry*. Springer-Verlag, New York, pp. 69-88; Stephenson, M.E., 1977. An approach to the identification of organic compounds hazardous to the environment and human health. *Ecotoxicol. Environ. Safety* 1, 407-425.

Gaussian 98 (Revision A.11.2), 2001. M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, V.G. Zakrzewski, J.A. Montgomery, Jr., R.E. Stratmann, J.C. Burant, S. Dapprich, J.M. Millam, A.D. Daniels, K.N. Kudin, M.C. Strain, O. Farkas, J. Tomasi, V. Barone, M. Cossi, R. Cammi, B. Mennucci, C. Poimelli, C. Adamo, S. Clifford, J. Ochterski, G.A. Petersson, P.Y. Ayala, Q. Cui, K. Morokuma, P. Salvador, J.J. Dannenberg, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J. Cioslowski, J.V. Ortiz, A.G. Baboul, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R. Gomperts, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, J.L. Andres, C. Gonzalez, M. Head-Gordon, E.S. Replogle, and J.A. Pople, Gaussian, Inc.: Pittsburgh, PA.

Geiss, K.T., Frazier, J.M., 2000. A novel in vitro system for exposures of cell cultures to volatile chemicals, *The Toxicologist* 54, 377.

Gute, B.D., Grunwald, G.D., Basak, S.C., 1999. Prediction of the dermal penetration of polycyclic aromatic hydrocarbons (PAHs): A hierarchical QSAR approach, SAR QSAR Environ. Res. 10, 1-15.

Hansch, C., Leo, A., 1995. Exploring QSAR: Fundamentals and Applications in Chemistry and Biology. American Chemical Society, Washington, D.C.

Kier, L.B., Hall, L.H., 1986. Molecular Connectivity in Structure-Activity Analysis, Research Studies Press, Letchworth, Hertfordshire, U.K.

Molconn-Z v3.50, 2000. Hall Associates Consulting: Quincy, MA.

Randic, M., 1975. J. Am. Chem. Soc. 97, 6609-6615.

SAS Institute, 1988. SAS/STAT User's Guide, Release 6.03, SAS Institute Inc.: SAS/STAT User's Guide, Release 6.03 Edition, Cary, NC.

Stewart, J.J.P., 1990. MOPAC Version 6.00. QCPE #455, Frank J. Seiler Research Laboratory, US Air Force Academy, CO.

Sybyl v6.4, 1997. Tripos Associates, Inc.: St. Louis, MO.

Trinajstić, N., 1992. Chemical Graph Theory. CRC Press.

Wiener, H., 1947. J. Am. Chem. Soc. 69, 17-20.

**Table 1.** Twenty Experimental Halocarbons for QSAR

Set	Formula	CAS #	Cat# (S/F)	Purity/Grade	Doses (mM)
P,M	CCl <sub>4</sub>	56-23-5	C5331(S)	99.9%	0.043 - 1.380
P,M	CBr <sub>4</sub>	558-13-4	C1,108-1(S)	99%	0.075 - 0.200
P,M	CHBrCl <sub>2</sub>	75-27-4	13,918-1(S)	98+%	0.500 - 10.000
P,M	CHBr <sub>2</sub> Cl	124-48-1	20,632-6(S)	98%	0.290 - 2.350
P,M	CBr <sub>2</sub> Cl <sub>2</sub>	594-18-3	32,995-9(S)	95%	0.100 - 1.000
P,M	CBrCl <sub>3</sub>	75-62-7	B8,225-1(S)	99%	0.310 - 2.500
M	CH <sub>2</sub> Br <sub>2</sub>	74-95-3	D4,168-6(S)	99%	2.500 - 25.000
M	CHCl <sub>3</sub>	67-66-3	BP1145-1(F)	99%	1.000 - 10.000
M	C <sub>2</sub> Cl <sub>4</sub>	127-18-4	O4586-4(F)	Certified Reagent	0.100 - 2.000
M	C <sub>2</sub> HCl <sub>3</sub>	79-01-6	25,642-0(S)	99.5%	0.277 - 2.660
M	1,2-C <sub>2</sub> H <sub>4</sub> Cl <sub>2</sub>	107-06-2	27,057-1(S)	99.8%	3.160 - 38.000
M	1,1,1-C <sub>2</sub> H <sub>3</sub> Cl <sub>3</sub>	71-55-6	23,557-1(S)	99%	1.080 - 6.480
M	1,1,2-C <sub>2</sub> H <sub>3</sub> Cl <sub>3</sub>	79-00-5	46,621-2(S)	97%	0.063 - 0.501
M	1,1,2-C <sub>2</sub> H <sub>3</sub> Br <sub>3</sub>	78-74-0	47,649-8(S)	99%	0.310 - 2.460
M	1,1,2,2-C <sub>2</sub> H <sub>2</sub> Cl <sub>4</sub>	79-34-5	18,543-4(S)	98+%	1.188 - 4.750
E	CH <sub>2</sub> Cl <sub>2</sub>	75-09-2	27,056-3(S)	99.9%	10.000 - 100.000
E	CHBr <sub>3</sub>	75-25-2	24,103-2(S)	99+%	0.143 - 4.580
E	CH <sub>2</sub> BrCl	74-97-5	13,526-7(S)	99%	10.000 - 100.00
E	1,2-C <sub>2</sub> H <sub>4</sub> BrCl	107-04-0	23,275-0(S)	98%	0.420 - 3.364
E	1,2-C <sub>2</sub> H <sub>4</sub> Br <sub>2</sub>	106-93-4	D887(S)	99%	0.070 - 8.630

P = preliminary model set; M = full model set; E = evaluation set; F = Fisher Scientific; S = Sigma/Adrich Company

**Table 2.** Symbols, definitions and classification of calculated molecular descriptors

<i>Topostructural (TS)</i>	
$I_D^W$	Information index for the magnitudes of distances between all possible pairs of vertices of a graph
$\bar{I}_D^W$	Mean information index for the magnitude of distance
W	Wiener index = half-sum of the off-diagonal elements of the distance matrix of a graph
$I^D$	Degree complexity
$H^V$	Graph vertex complexity
$H^D$	Graph distance complexity
$\overline{IC}$	Information content of the distance matrix partitioned by frequency of occurrences of distance h
$M_1$	A Zagreb group parameter = sum of square of degree over all vertices
$M_2$	A Zagreb group parameter = sum of cross-product of degrees over all neighboring (connected) vertices
${}^h\chi$	Path connectivity index of order h = 0-6
${}^h\chi_C$	Cluster connectivity index of order h = 3-6
${}^h\chi_{PC}$	Path-cluster connectivity index of order h = 4-5
$P_h$	Number of paths of length h = 0, 2-7
J	Balaban's J index based on topological distance
$DN^2S_y$	Triplet index from distance matrix, square of graph order (# of non-H atoms), and distance sum; operation y = 1-5
$DN^21_y$	Triplet index from distance matrix, square of graph order, and number 1; operation y = 1-4
$AS1_y$	Triplet index from adjacency matrix, distance sum, and number 1; operation y = 1-5
$DS1_y$	Triplet index from distance matrix, distance sum, and number 1; operation y = 1-5
$ASN_y$	Triplet index from adjacency matrix, distance sum, and graph order; operation y = 1-5
$DSN_y$	Triplet index from distance matrix, distance sum, and graph order; operation y = 1-5
$DN^2N_y$	Triplet index from distance matrix, square of graph order, and graph order; operation y = 1-5
$ANS_y$	Triplet index from adjacency matrix, graph order, and distance sum; operation y = 1-2
$AN1_y$	Triplet index from adjacency matrix, graph order, and number 1; operation y = 1-5
$ANN_y$	Triplet index from adjacency matrix, graph order, and graph order again; operation y = 1-5
$ASV_y$	Triplet index from adjacency matrix, distance sum, and vertex degree; operation y = 1-5
$DSV_y$	Triplet index from distance matrix, distance sum, and vertex degree; operation y = 1-2



ANV <sub>y</sub>	Triplet index from adjacency matrix, graph order, and vertex degree; operation y = 1-5
<i>Topochemical (TC)</i>	
O	Order of neighborhood when IC <sub>r</sub> reaches its maximum value for the hydrogen-filled graph
O <sub>orb</sub>	Order of neighborhood when IC <sub>r</sub> reaches its maximum value for the hydrogen-suppressed graph
I <sub>ORB</sub>	Information content or complexity of the hydrogen-suppressed graph at its maximum neighborhood of vertices
IC <sub>r</sub>	Mean information content or complexity of a graph based on the r <sup>th</sup> (r = 0-5) order neighborhood of vertices in a hydrogen-filled graph
SIC <sub>r</sub>	Structural information content for r <sup>th</sup> (r = 0-5) order neighborhood of vertices in a hydrogen-filled graph
CIC <sub>r</sub>	Complementary information content for r <sup>th</sup> (r = 0-5) order neighborhood of vertices in a hydrogen-filled graph
$\chi^b_h$	Bond path connectivity index of order h = 0-4
$\chi^b_C$	Bond cluster connectivity index of order h = 3 & 5
$\chi^b_{PC}$	Bond path-cluster connectivity index of order h = 4-5
$\chi^v_h$	Valence path connectivity index of order h = 0-6
$\chi^v_C$	Valence cluster connectivity index of order h = 3-5
$\chi^v_{PC}$	Valence path-cluster connectivity index of order h = 4-5
J <sup>B</sup>	Balaban's J index based on bond types
J <sup>X</sup>	Balaban's J index based on relative electronegativities
J <sup>Y</sup>	Balaban's J index based on relative covalent radii
AZV <sub>y</sub>	Triplet index from adjacency matrix, atomic number, and vertex degree; operation y = 1-5
AZS <sub>y</sub>	Triplet index from adjacency matrix, atomic number, and distance sum; operation y = 1-5
ASZ <sub>y</sub>	Triplet index from adjacency matrix, distance sum, and atomic number; operation y = 1-2
AZN <sub>y</sub>	Triplet index from adjacency matrix, atomic number, and graph order; operation y = 1-5
ANZ <sub>y</sub>	Triplet index from adjacency matrix, graph order, and atomic number; operation y = 1-2
DSZ <sub>y</sub>	Triplet index from distance matrix, distance sum, and atomic number; operation y = 1-2
DN <sup>2</sup> Z <sub>y</sub>	Triplet index from distance matrix, square of graph order, and atomic number; operation y = 1-2
nelem	Number of elements in a molecule
fw	Molecular weight
totop	Total Topological Index t
sumI	Sum of the intrinsic state values I
sumdell	Sum of delta-I values
tets2	Total topological state index based on electrotopological state indices
phia	Flexibility index (kp1 * kp2/nvx)

$I_D^C$	Bonchev-Trinajstić information index
$\bar{I}_D^C$	Mean Bonchev-Trinajstić information index
knotp	Difference of chi-cluster-3 and path/cluster-4
knotpv	Valence difference of chi-cluster-3 and path/cluster-4
nclass	Number of classes of topologically (symmetry) equivalent graph vertices
numwHBd	Number of weak hydrogen bond donors
numHBa	Number of hydrogen bond acceptors
SHCsats	E-State of C $sp^3$ bonded to other saturated C atoms
SHCsatu	E-State of C $sp^3$ bonded to unsaturated C atoms
SHvin	E-State of C atoms in the vinyl group, =CH-
SHtvin	E-State of C atoms in the terminal vinyl group, =CH <sub>2</sub>
SHwHBd	Weak hydrogen bond donor index, sum of C-H Hydrogen E-State values for hydrogen atoms on a C to which a F and/or Cl are also bonded
SHHBa	Hydrogen bond acceptor index, sum of the E-State values for –OH, =NH, –NH <sub>2</sub> , –NH–, >N–, –O–, –S–, along with –F and –Cl
Qv	General Polarity descriptor Electrotopological State index values for atoms types: SHother, SHCHnX, Hmax Gmax, Hmin, SsCH <sub>3</sub> , SdCH <sub>2</sub> , SssCH <sub>2</sub> , SdsCH, SdssC, SsF, SsCl, SsBr
<b><i>Geometrical / Shape (3D)</i></b>	
kp0	Kappa zero
kp2-kp3	Kappa simple indices
ka1-ka3	Kappa alpha indices
V <sub>w</sub>	Van der Waals volume
<sup>3D</sup> W <sub>H</sub>	3D Wiener number based on the hydrogen-filled geometric distance matrix
<sup>3D</sup> W	3D Wiener number based on the hydrogen-suppressed geometric distance matrix
<b><i>Semi-empirical Quantum Chemical (QC)</i></b>	
E <sub>HOMO</sub>	Energy of the highest occupied molecular orbital
E <sub>HOMO-1</sub>	Energy of the second highest occupied molecular orbital
E <sub>LUMO</sub>	Energy of the lowest unoccupied molecular orbital
E <sub>LUMO+1</sub>	Energy of the second lowest unoccupied molecular orbital
GAP <sub>H-L</sub>	HOMO-LUMO gap
ΔH <sub>f</sub>	Heat of formation
μ	Dipole moment
<b><i>Ab initio Quantum Chemical (QC)</i></b>	
E <sub>HOMO</sub>	HOMO-LUMO gap
E <sub>HOMO-1</sub>	
E <sub>LUMO</sub>	
E <sub>LUMO+1</sub>	
GAP <sub>H-L</sub>	
μ	

**Table 3.** Statistical results for 5 models on 6 biological properties for measuring cellular toxicity of twenty halocarbons.

<b>Biological Property</b>		<b>Topological</b>	<b>AM1</b>	<b>STO-3G</b>	<b>TI + AM1</b>	<b>TI + STO-3G</b>
MTT	$R^2_{cv}$	0.3178	0.3027	0.2653	0.5014	<i>same as</i>
	<i># indep. var.</i>	3	4	2	3	<i>topological</i>
	$F$	8.68	6.14	7.61	13.15	<i>model</i>
	<i>s.e.</i>	1.16	1.20	1.33	1.01	
LDH	$R^2_{cv}$	0.6776	0.4814	0.4357	0.6875	<i>same as</i>
	<i># indep. var.</i>	3	4	3	4	<i>topological</i>
	$F$	19.44	8.31	9.74	16.11	<i>model</i>
	<i>s.e.</i>	0.85	1.05	1.09	0.82	
ROS	$R^2_{cv}$	0.6976	0.1060	0.0757	<i>same as</i>	0.6849
	<i># indep. var.</i>	4	2	3	<i>topological</i>	3
	$F$	17.01	5.06	3.79	<i>model</i>	20.88
	<i>s.e.</i>	0.90	1.57	1.56		0.92
TBARS	$R^2_{cv}$	0.5914	0.1738	0.1790	<i>same as</i>	0.6173
	<i># indep. var.</i>	3	3	2	<i>topological</i>	3
	$F$	15.19	3.81	5.68	<i>model</i>	15.23
	<i>s.e.</i>	0.85	1.28	1.26		0.85
Thiols	$R^2_{cv}$	0.6074	0.1703	0.0705	<i>same as</i>	0.6299
	<i># indep. var.</i>	4	3	3	<i>topological</i>	4
	$F$	13.22	3.96	3.43	<i>model</i>	14.78
	<i>s.e.</i>	0.94	1.47	1.51		0.90
Catalase	$R^2_{cv}$	0.6481	0.4185	0.3269	<i>same as</i>	0.6442
	<i># indep. var.</i>	4	2	2	<i>topological</i>	4
	$F$	16.71	9.65	8.26	<i>model</i>	16.87
	<i>s.e.</i>	0.83	1.24	1.29		0.82

## **Figure Captions.**

### **Figure 1. Cell culture protocol.**

t = -4h: Time that the cells were first plated in attachment media with dexamethasone

t = 0 h: Time at which attachment media replaced with normal CHEEs culture media

t = 20 h: Time at which volatile chemical exposures began

t = 24 h: End of the volatile chemical exposure and beginning of biochemical assays

